



#4

PATENT APPLICATION

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of

Katsumi AMANO, et al.

Appln. No.: 09/684,331

Group Art Unit: 2644

Confirmation No.: NOT YET ASSIGNED

Examiner: NOT YET ASSIGNED

Filed: October 10, 2000

For: SPEECH SYNTHESIS METHOD

SUBMISSION OF PRIORITY DOCUMENT

Commissioner for Patents
Washington, D.C. 20231

Sir:

Submitted herewith is a certified copy of the priority document on which a claim to priority was made under 35 U.S.C. § 119. The Examiner is respectfully requested to acknowledge receipt of said priority document.

Respectfully submitted,

SUGHRUE, MION, ZINN,
MACPEAK & SEAS, PLLC
2100 Pennsylvania Avenue, N.W.
Washington, D.C. 20037-3213
Telephone: (202) 293-7060
Facsimile: (202) 293-7860

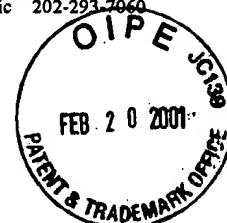
Paul J. Niles Reg. 33102
for / Darryl Mexic
Registration No. 23,063

Enclosures: JP 11-294357

DM/alb

Date: February 20, 2001

日 本 国 特 許 庁
PATENT OFFICE
JAPANESE GOVERNMENT



別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日
Date of Application:

1999年10月15日

出 願 番 号
Application Number:

平成11年特許願第294357号

出 願 人
Applicant (s):

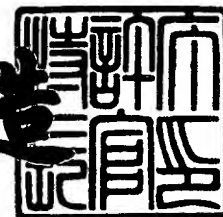
パイオニア株式会社

CERTIFIED COPY OF
PRIORITY DOCUMENT

2000年 8月11日

特許庁長官
Commissioner,
Patent Office

及 川 耕 造



出証番号 出証特2000-3063374

【書類名】 特許願

【整理番号】 54P0136

【提出日】 平成11年10月15日

【あて先】 特許庁長官殿

【国際特許分類】 G10L 5/04

【発明の名称】 音声合成方法

【請求項の数】 4

【発明者】

 【住所又は居所】 埼玉県鶴ヶ島市富士見6丁目1番1号 パイオニア株式会社 総合研究所内

 【氏名】 天野 克己

【発明者】

 【住所又は居所】 埼玉県鶴ヶ島市富士見6丁目1番1号 パイオニア株式会社 総合研究所内

 【氏名】 張 子青

【発明者】

 【住所又は居所】 埼玉県鶴ヶ島市富士見6丁目1番1号 パイオニア株式会社 総合研究所内

 【氏名】 外山 聡一

【発明者】

 【住所又は居所】 埼玉県鶴ヶ島市富士見6丁目1番1号 パイオニア株式会社 総合研究所内

 【氏名】 石原 博幸

【特許出願人】

 【識別番号】 000005016

 【氏名又は名称】 パイオニア株式会社

【代理人】

 【識別番号】 100079119

 【弁理士】

【氏名又は名称】 藤村 元彦

【手数料の表示】

【予納台帳番号】 016469

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 9006557

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 音声合成方法

【特許請求の範囲】

【請求項 1】 周波数信号を発生する音源と、音素系列中の各音素に対応した線形予測係数に応じたフィルタ特性にて前記周波数信号をフィルタリング処理することにより音声波形信号を得る声道フィルタと、を備えた音声合成装置における音声合成方法であって、

前記音素を各々が所定期間長からなる複数のフレームに分割し、

前記音素中の前記フレーム毎にフレーム内に存在する音声サンプルの 2 乗和をフレームパワー値として求め、

前記音素中の先頭フレーム及び最後尾フレーム各々での前記フレームパワー値を所定値に規格化した際における n 番目の前記フレームでのフレームパワー値を規格化フレームパワー値として求め、

前記周波数信号における 1 フレーム期間内での信号レベルの 2 乗和をフレームパワー補正值として求め、

前記規格化フレームパワー値及び前記フレームパワー補正值を変数とする関数によって定まるレベルの音声包絡信号に応じて前記音声波形信号の振幅レベルを調整することを特徴とする音声合成方法。

【請求項 2】 前記 n 番目の前記フレームに対応した前記線形予測係数に基づいて電力周波数特性を求め、前記電力周波数特性から所定周波数間隔毎に採取した電力値の平均値を平均フレームパワー値として求め、

前記規格化フレームパワー値、前記フレームパワー補正值、及び前記平均フレームパワー値を変数とする関数によって定まるレベルの音声包絡信号に応じて前記音声波形信号の振幅レベルを調整することを特徴とする請求項 1 記載の音声合成方法。

【請求項 3】 前記規格化フレームパワー値を P_n 、前記フレームパワー補正值を G_s 、前記平均フレームパワー値を G_f と表した場合に前記関数は、

$$\sqrt{\{P_n / (G_s \cdot G_f)\}}$$

であることを特徴とする請求項 2 記載の音声合成方法。

【請求項 4】 前記周波数信号とは、有声音を担うインパルス信号及び無声音を担う雑音信号からなることを特徴とする請求項 1 記載の音声合成方法。

【発明の詳細な説明】

【 0 0 0 1 】

【発明の属する技術分野】

本発明は、人工的に音声波形信号を生成する音声合成(voice synthesis)に関する。

【 0 0 0 2 】

【背景技術】

自然音声による音声波形は、音素、すなわち 1 つの母音(以下、V と称する)と、1 つの子音(以下、C と称する)とが、"CV"、"CVC"、又は"VCV"の如く連続してなる基本単位を連結することによって表すことが出来る。

従って、このような音素各々を予めデータ化(音素データ)して登録しておき、その中から会話に対応した音素データのみを順次読み出し、各音素データに対応した音を順次発生させて行けば、この会話を合成音声によって実施させることが可能となる。

【 0 0 0 3 】

この際、上記音素データによるデータベースを構築するには、先ず、所定の文書を人に読み上げて貰い、その音声を一旦録音する。次に、この録音した音声を再生して得られた再生音声信号中から上記音素の切り出しを行い、この音素を示す各種のデータを音素データとして登録して行くのである。そして、音声合成時には、各音素データを接続して一連の音声として出力して行く。

【 0 0 0 4 】

しかしながら、接続する各音素は別々に録音された音声から切り出したものである為、発声時の音声パワーにバラツキがあり、これを単に連結しただけでは不自然な合成音声しか得られないという問題が発生した。

【 0 0 0 5 】

【発明が解決しようとする課題】

本発明は、自然な合成音声が得られる音声合方法を提供することを目的とする

【 0 0 0 6 】

【課題を解決するための手段】

本発明による音声合成方法は、周波数信号を発生する音源と、音素系列中の各音素に対応した線形予測係数に応じたフィルタ特性にて前記周波数信号をフィルタリング処理することにより音声波形信号を得る声道フィルタと、を備えた音声合成装置における音声合成方法であって、前記音素を各々が所定期間長からなる複数のフレームに分割し、前記音素中の前記フレーム毎にフレーム内に存在する音声サンプルの 2 乗和をフレームパワー値として求め、前記音素中の先頭フレーム及び最後尾フレーム各々での前記フレームパワー値を所定値に規格化した際における n 番目の前記フレームでのフレームパワー値を規格化フレームパワー値として求め、前記周波数信号における 1 フレーム期間内での信号レベルの 2 乗和をフレームパワー補正值として求めて、前記規格化フレームパワー値及び前記フレームパワー補正值を変数とする関数によって定まるレベルの音声包絡信号に応じて前記音声波形信号の振幅レベルを調整する。

【 0 0 0 7 】

【発明の実施の形態】

図 1 は、本発明による音声合成方法に基づいて音声合成(voice synthesis)を行って、所望の文書(テキスト)の読み上げを行わせるテキスト音声合成装置の構成を示す図である。

図 1 において、音源モジュール 2 3 は、上記音声合成制御回路 2 2 から供給されたピッチ周波数指定信号 K に応じた周波数のインパルス信号を発生するパルス発生器 2 3 1 と、無声音を担う雑音信号を発生する雑音発生器 2 3 2 とを備えている。音源モジュール 2 3 は、上記インパルス信号及び雑音信号の内から、上記音声合成制御回路 2 2 から供給された音源選択信号 S_v にて示される方を択一的に選択し、これを周波数信号 Q として声道フィルタ 2 4 に供給する。

【 0 0 0 8 】

声道フィルタ 2 4 は、例えば F I R (Finite Impulse Response) デジタルフィルタ等からなり、上記音声合成制御回路 2 2 から供給された線形予測符号信号

LPに応じたフィルタ係数にて、上記音源モジュール 2 3 から供給された周波数信号 Q に対してフィルタリング処理を施して音声波形信号 V_F を得る。

振幅調整回路 2 5 は、かかる音声波形信号 V_F の振幅を、音声合成制御回路 2 2 から供給された音声包絡信号 V_m に応じた振幅に調整して得られた振幅調整音声波形信号 V_{AUD} をスピーカ 2 6 に供給する。スピーカ 2 6 は、かかる振幅調整音声波形信号 V_{AUD} に応じた音響出力を行う。この際、スピーカ 2 6 からは、以下に説明するが如き、入力されたテキスト信号に基づく読み上げ音声が出力される。

【 0 0 0 9 】

テキスト解析回路 2 1 は、入力されたテキスト信号に基づく文字列に、各言語特有のアクセント及びフレーズ等の情報を織り込んだ中間言語文字列情報を生成し、この情報に対応した中間言語文字列信号 CL を音声合成制御回路 2 2 に供給する。

音声合成制御回路 2 2 には、音素データメモリ 2 0、RAM(Random Access Memory) 2 7、及びROM(Read Only Memory) 2 8 が接続されている。

【 0 0 1 0 】

音素データメモリ 2 0 には、人間の実際の音声から採取した各種音素に対応した音素データ、及び音声合成時に用いる音声合成パラメータ(規格化フレームパワー値、平均フレームパワー値)が予め格納されている。

以下に、かかる音素データメモリ 2 0 に格納される上記音素データ及び音声合成パラメータの生成方法について述べる。

【 0 0 1 1 】

図 2 は、かかる音声合成パラメータを生成する際のシステム構成を示す図である。

図 2 において、音声レコーダ 3 2 は、マイクロフォン 3 1 によって集音した人間の音声を一旦録音する。音声レコーダ 3 2 は、この録音した音声を再生して得た再生音声信号を音素データ生成装置 3 0 に供給する。

【 0 0 1 2 】

音素データ生成装置 3 0 は、上記音声レコーダ 3 2 から供給された再生音声信

号をサンプリングして得た音声サンプルを順次、メモリ 33 内の所定領域に記憶した後、先ず、図 3 に示されるが如き音素データ生成行程を実行する。

図 3 において、先ず、音素データ生成装置 30 は、上記メモリ 33 に記憶した音声サンプルを順次読み出し、その音声波形に基づき、かかる音声サンプルの系列を例えば”VCV”なる音素に区切る(ステップ S1)。

【0013】

例えば、”目的地に”なる音声は、

mo/oku/ute/eki/iti/ini/i

”催し物の”なる音声は、

mo/oyo/osi/imo/ono/ono/o

”最寄りの”なる音声は、

mo/oyo/ori/ino/o

”目標の”なる音声は、

mo/oku/uhyo/ono/o

なる音素に夫々区切られる。

【0014】

次に、音素データ生成装置 30 は、この切り出した音素各々を、所定長、例えば 10 [msec] 毎のフレームに分割し(ステップ S2)、分割したフレームの各々に、そのフレームが属する音素の名前、及びこの音素のフレーム長、並びにフレーム番号等の管理情報を付加したものを、メモリ 33 の所定領域に記憶する(ステップ S3)。次に、音素データ生成装置 30 は、各音素の波形に対して、上記フレーム毎に線形予測符号分析、いわゆる LPC (linear predictive coding) 分析を行って、例えば 15 次数分の線形予測符号係数(以下、LPC 係数と称する)を求め、これを図 4 に示されるが如き、メモリ 33 のメモリ領域 1 内に記憶する(ステップ S4)。尚、このステップ S4 において求めた LPC 係数とは、声道フィルタ 24 のフィルタ係数に相当する、いわゆる音声スペクトル包絡パラメータである。次に、音素データ生成装置 30 は、上記ステップ S4 においてメモリ 33 のメモリ領域 1 内に記憶した LPC 係数を読み出し、これを音素データとして出力する(ステップ S5)。かかる音素データが上記音素データメモリ 20 に格納

される。

【 0 0 1 5 】

次に、音素データ生成装置 3 0 は、メモリ 3 3 内に記憶した音素の各々に対して、図 5 に示されるが如き音声合成パラメータ算出行程を実行する。

図 5 において、音素データ生成装置 3 0 は、先ず、処理対象となる 1 音素(以下、対象音素と称する)中の各フレーム毎にそのフレーム内に存在する全ての音声サンプル値の二乗和を算出することにより、かかるフレーム内での音声パワーを求め、これをフレームパワー P_C として図 4 に示されるが如きメモリ 3 3 のメモリ領域 2 内に記憶する(ステップ S 1 2)。

【 0 0 1 6 】

次に、音素データ生成装置 3 0 は、先頭のフレーム番号を示す"0"を内蔵レジスタ n (図示せぬ)に記憶する(ステップ S 1 3)。次に、音素データ生成装置 3 0 は、上記内蔵レジスタ n に記憶されているフレーム番号によって示されるフレーム n の、この対象音素内での相対位置、すなわち、

$$r = (n - 1) / N$$

r : 相対位置

N : 対象音素中に存在する全フレーム数

を求める(ステップ S 1 4)。

【 0 0 1 7 】

次に、音素データ生成装置 3 0 は、上記内蔵レジスタ n によって示されるフレーム n でのフレームパワー P_C を図 4 に示されるが如きメモリ 3 3 のメモリ領域 2 から読み出す(ステップ S 1 5)。更に、音素データ生成装置 3 0 は、かかるメモリ領域 2 に記憶されているフレームパワー P_C の中から、上記対象音素の先頭フレーム及び最終フレーム各々に対応したものを夫々、先頭フレームパワー P_a 及び最終フレームパワー P_b として読み出す(ステップ S 1 6)。

【 0 0 1 8 】

次に、音素データ生成装置 3 0 は、これら先頭フレームパワー P_a 及び最終フレームパワー P_b と、上記ステップ S 1 5 において読み出したフレームパワー P_C と、上記相対位置 r を用いて、

$$P_n = P_c / [(1-r) \cdot P_a + r \cdot P_b]$$

なる演算を実施することにより、上記内蔵レジスタ n によって示されるフレーム n での規格化フレームパワー P_n を求め、これを図 4 に示されるが如き、メモリ 33 のメモリ領域 3 に記憶する(ステップ S 17)。

【0019】

すなわち、この対象音素の最後尾のフレームにおけるフレームパワー P_c を "1" とした場合に、上記フレーム n でのフレームパワー値がどのくらいになるかを求めるのである。

次に、音素データ生成装置 30 は、上記内蔵レジスタ n によって示されるフレーム n に対応した LPC 係数を図 4 に示されるが如きメモリ 33 のメモリ領域 1 から読み出し、かかる LPC 係数に基づいてこのフレーム n 内での電力周波数特性を求める(ステップ S 18)。次に、音素データ生成装置 30 は、かかる電力周波数特性から所定周波数間隔毎にその電力値を採取し、この電力値の平均を平均フレームパワー G_f として図 4 に示されるが如きメモリ 33 のメモリ領域 4 に記憶する(ステップ S 19)。

【0020】

次に、音素データ生成装置 30 は、上記内蔵レジスタ n に記憶されているフレーム番号 n に "1" を加算したものを新たなフレーム番号 n として、この内蔵レジスタ n に上書き記憶する(ステップ S 20)。次に、音素データ生成装置 30 は、内蔵レジスタ n に記憶されているフレーム番号が $(N-1)$ と等しくなったか否かの判定を行う(ステップ S 21)。

【0021】

かかるステップ S 21 において、内蔵レジスタ n に記憶されているフレーム番号が $(N-1)$ と等しくないと判定された場合、音素データ生成装置 30 は、上記ステップ S 14 の実行に戻って前述した如き動作を繰り返し実行する。かかる動作により、対象音素中の先頭のフレーム $\sim (N-1)$ 番目のフレーム各々に対する規格化フレームパワー P_n 及び平均フレームパワー G_f が図 4 に示されるが如きメモリ領域 3 及び 4 に順次記憶されて行くのである。

【0022】

一方、上記ステップ S 2 1 において、内蔵レジスタ n に記憶されているフレーム番号が $(N - 1)$ と等しいと判定された場合、音素データ生成装置 3 0 は、図 4 に示されるが如く、メモリ 3 3 のメモリ領域 3 及び 4 に記憶されている規格化フレームパワー P_n 及び平均フレームパワー G_f を夫々読み出して、これを出力する(ステップ S 2 3)。これら規格化フレームパワー P_n 及び平均フレームパワー G_f が音声合成パラメータとして、上記音素データメモリ 2 0 に格納される。

【 0 0 2 3 】

すなわち、図 3 に示される音素データ生成行程によって求められた各音素データと、図 5 に示される音声合成パラメータ算出行程によって求められた規格化フレームパワー P_n 及び平均フレームパワー G_f とが互いに対応付けられて、音素データメモリ 2 0 に格納されるのである。

図 1 に示される音声合成制御回路 2 2 は、ROM 2 8 に記憶されているソフトウェアに従って、上記テキスト解析回路 2 1 から供給されてきた中間言語文字列信号 CL に応じた音素データ及び音声合成パラメータを取り込んで、以下に説明するが如き音声合成制御を行う。

【 0 0 2 4 】

音声合成制御回路 2 2 は、先ず、上記中間言語文字列信号 CL を "VCV" なる音素に区切り、各音素に対応した音素データを音素データメモリ 2 0 から順次読み出して取り込む。更に、音声合成制御回路 2 2 は、ピッチ周波数を指定するピッチ周波数指定信号 K を音源モジュール 2 3 に供給する。次に、音声合成制御回路 2 2 は、音素データメモリ 2 0 から読み出した順に、音素データ各々に応じた音声合成処理を実行する。

【 0 0 2 5 】

図 6 は、かかる音声合成処理を司る音声合成制御ルーチンを示す図である。

図 6 において、先ず、音声合成制御回路 2 2 は、上述の如く取り込んだ順に処理対象とすべき 1 音素データ(以下、対象音素データと称する)を選出し、この音素データにおける先頭のフレーム番号を示す "0" を内蔵レジスタ n (図示せぬ)に記憶する(ステップ S 1 0 1)。次に、音声合成制御回路 2 2 は、かかる対象音素データに対応した音素が有声音であるのか、あるいは無声音であるのかを示す音

源選択信号 S_v を音源モジュール 23 に供給する (ステップ S102)。これにより、音源モジュール 23 は、上記ピッチ周波数指定信号 K によって示される周波数を有するインパルス信号、又は雑音信号を周波数信号 Q として出力する。

【0026】

次に、音声合成制御回路 22 は、上記音源モジュール 23 から供給された周波数信号 Q を所定サンプリングタイミング毎にサンプリングして取り込み、そのサンプル値各々の 1 フレーム内での 2 乗和をフレームパワー補正值 G_s として求め、これを内蔵レジスタ G (図示せぬ) に記憶する (ステップ S103)。次に、音声合成制御回路 22 は、上記対象音素データ中における上記内蔵レジスタ n にて示されるフレーム n に対応した LPC 係数を線形予測符号信号 LP として声道フィルタ 24 に供給する (ステップ S104)。次に、音声合成制御回路 22 は、上記対象音素データ中の上記内蔵レジスタ n によって示されるフレーム n に対応した規格化フレームパワー P_n 及び平均フレームパワー G_f を音素データメモリ 20 から読み出す (ステップ S105)。次に、音声合成制御回路 22 は、上記規格化フレームパワー P_n 、平均フレームパワー G_f 及び上記内蔵レジスタ G に記憶した上記フレームパワー補正值 G_s を用いた以下の演算により、音声包絡信号 V_m を求め、これを振幅調整回路 25 に供給する (ステップ S106)。

【0027】

$$V_m = \sqrt{\{P_n / (G_s \cdot G_f)\}}$$

かかるステップ S106 の実行により、振幅調整回路 25 は、声道フィルタ 24 から供給された音声波形信号 V_F の振幅を上記音声包絡信号 V_m に応じた振幅に調整する。この振幅調整によれば、各音素の連結部分が常に所定レベルに保たれるようになるので、音素間の繋がりが滑らかになり自然な合成音声を得られるようになる。

【0028】

次に、音声合成制御回路 22 は、内蔵レジスタ n に記憶されているフレーム番号 n と、上記対象音素データ中の全フレーム数 N よりも 1 だけ小なる値、すなわち $(N-1)$ とが一致しているか否かを判定する (ステップ S107)。かかるステップ S107 において、両者が一致していない判定された場合、音声合成制御回

路 22 は、上記内蔵レジスタ n に記憶されているフレーム番号に "1" を加算したものを新たなフレーム番号として、この内蔵レジスタ n に上書き記憶する (ステップ S108)。かかるステップ S108 の実行後、音声合成制御回路 22 は、上記ステップ S103 の実行に戻って、前述した如き動作を繰り返し実行する。

【0029】

一方、上記ステップ S107 において、内蔵レジスタ n に記憶されているフレーム番号 n と、 $(N-1)$ とが一致していないと判定された場合、音声合成制御回路 22 は、上記ステップ S101 の実行に戻って、次の音素データに対する音声合成処理を同様に実施する。

【0030】

【発明の効果】

以上、詳述した如く、本発明による音声合成方法においては、音素中の先頭フレーム及び最後尾フレーム各々でのフレームパワー値を所定値に規格化した際における n 番目のフレームでの規格化フレームパワー値を求め、この n 番目のフレームにおける電力周波数特性から所定周波数間隔毎に採取した電力値の平均を平均フレームパワー値として求め、更に、音源からの周波数信号の 1 フレーム期間内での信号レベルの 2 乗和をフレームパワー補正值として夫々求め、これら規格化フレームパワー値、フレームパワー補正值、及び平均フレームパワー値を変数とする関数によって定まるレベルの音声包絡信号に応じて、声道フィルタから出力された音声波形信号の振幅レベルを調整するようにしている。

【0031】

従って、本発明によれば、合成音声波形を大幅に変形させることなく、各音素の先頭部及び後尾部のレベルが常に所定レベルに保たれるようになるので、各音素が滑らかに繋がるようになり、自然な合成音声を出力させることが出来る。

【図面の簡単な説明】

【図 1】

本発明による音声合成装置の構成を示す図である。

【図 2】

音素データ及び音声合成パラメータを生成する際のシステム構成を示す図であ

る。

【図 3】

音素データの生成行程を示す図である。

【図 4】

メモリ 3 3 のメモリマップを示す図である。

【図 5】

音声合成パラメータ算出行程を示す図である。

【図 6】

本発明の音声合成方法に基づく音声合成制御ルーチンを示す図である。

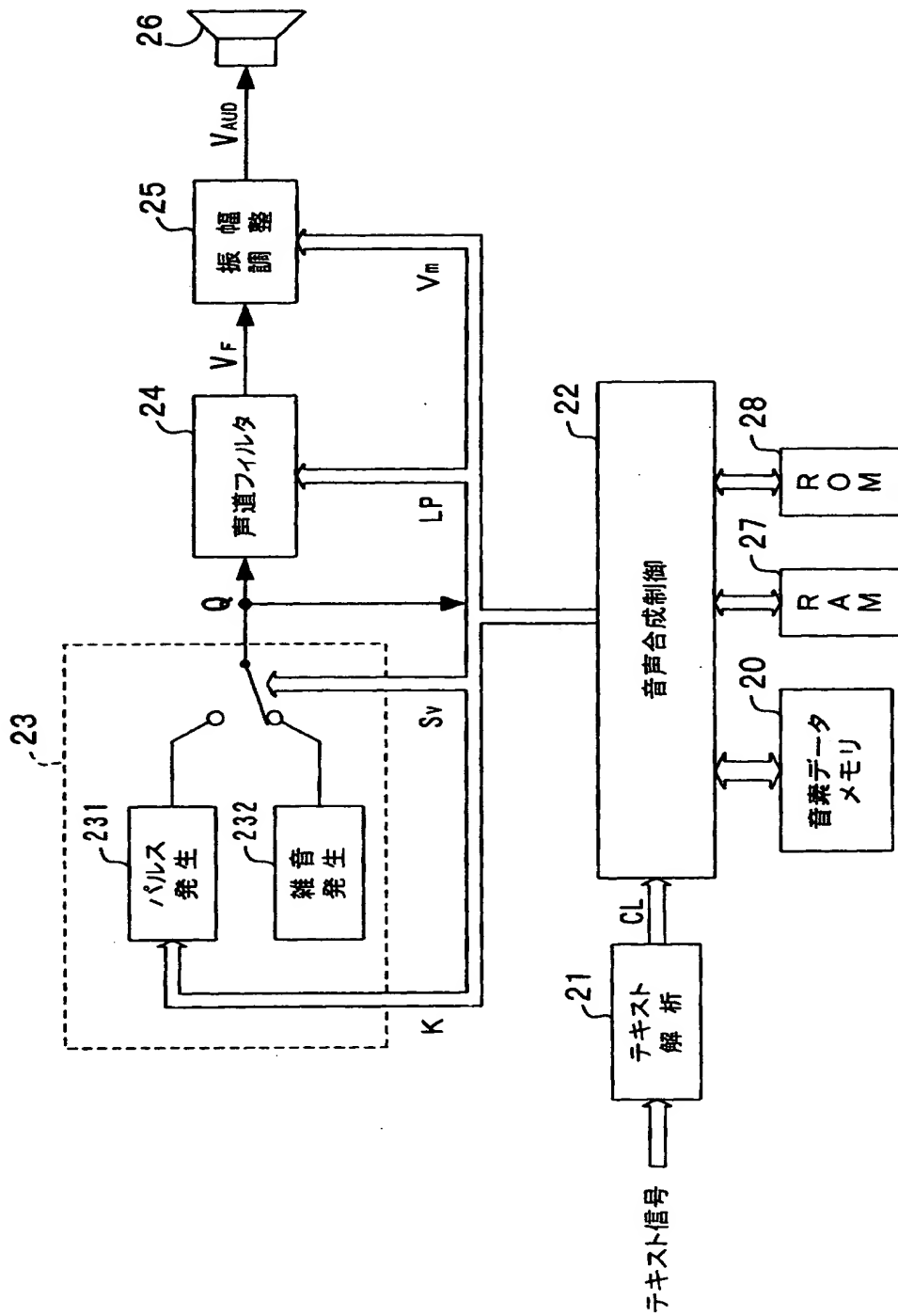
【主要部分の符号の説明】

- 2 0 音素データメモリ
- 2 2 音声合成制御回路
- 3 0 音素データ生成装置
- 3 3 メモリ
- 2 3 音源モジュール
- 2 4 声道フィルタ
- 2 5 振幅調整回路

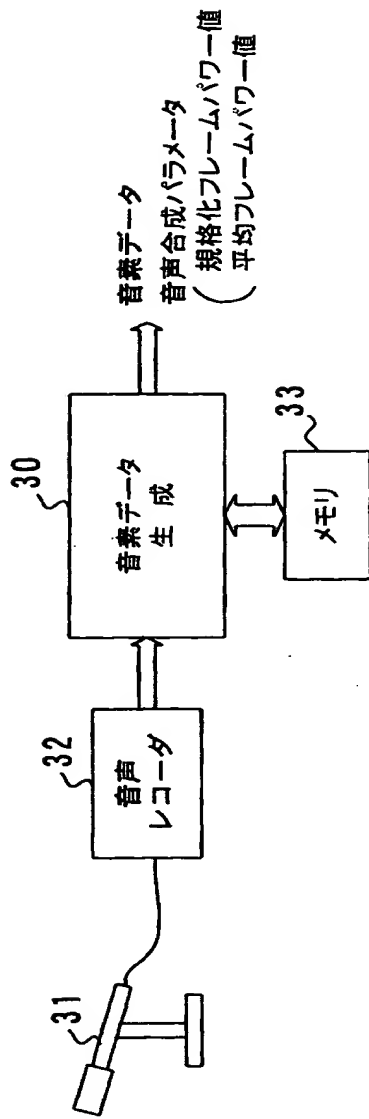
【書類名】

図面

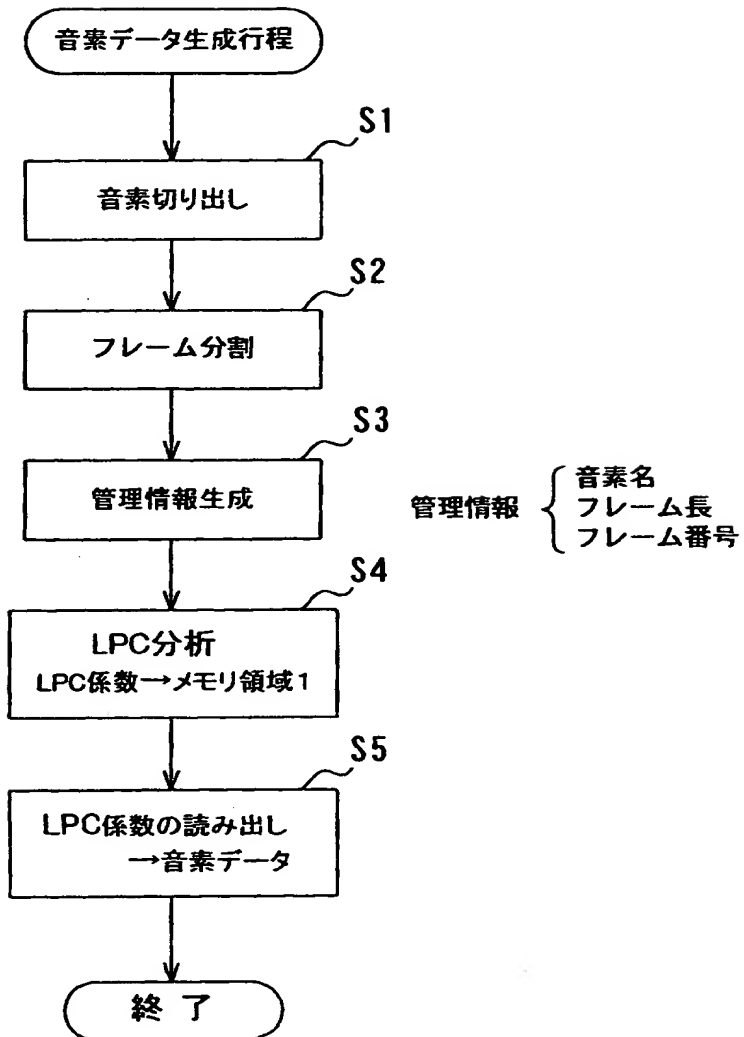
【図 1】



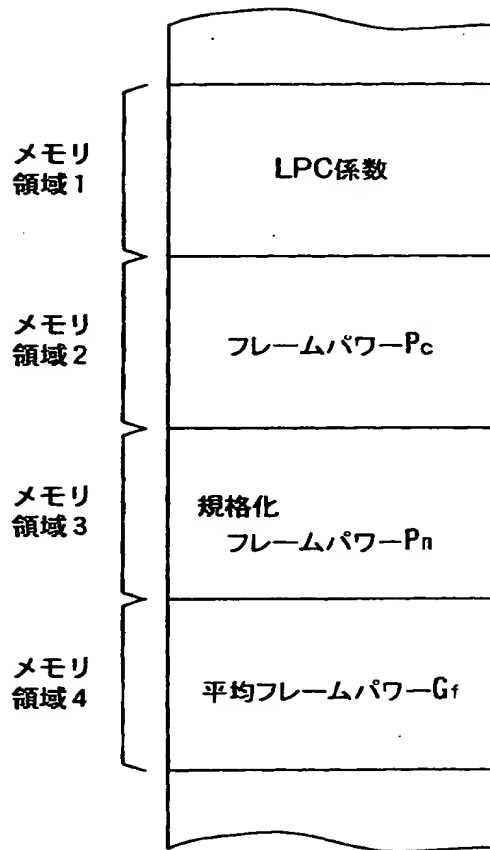
【図 2】



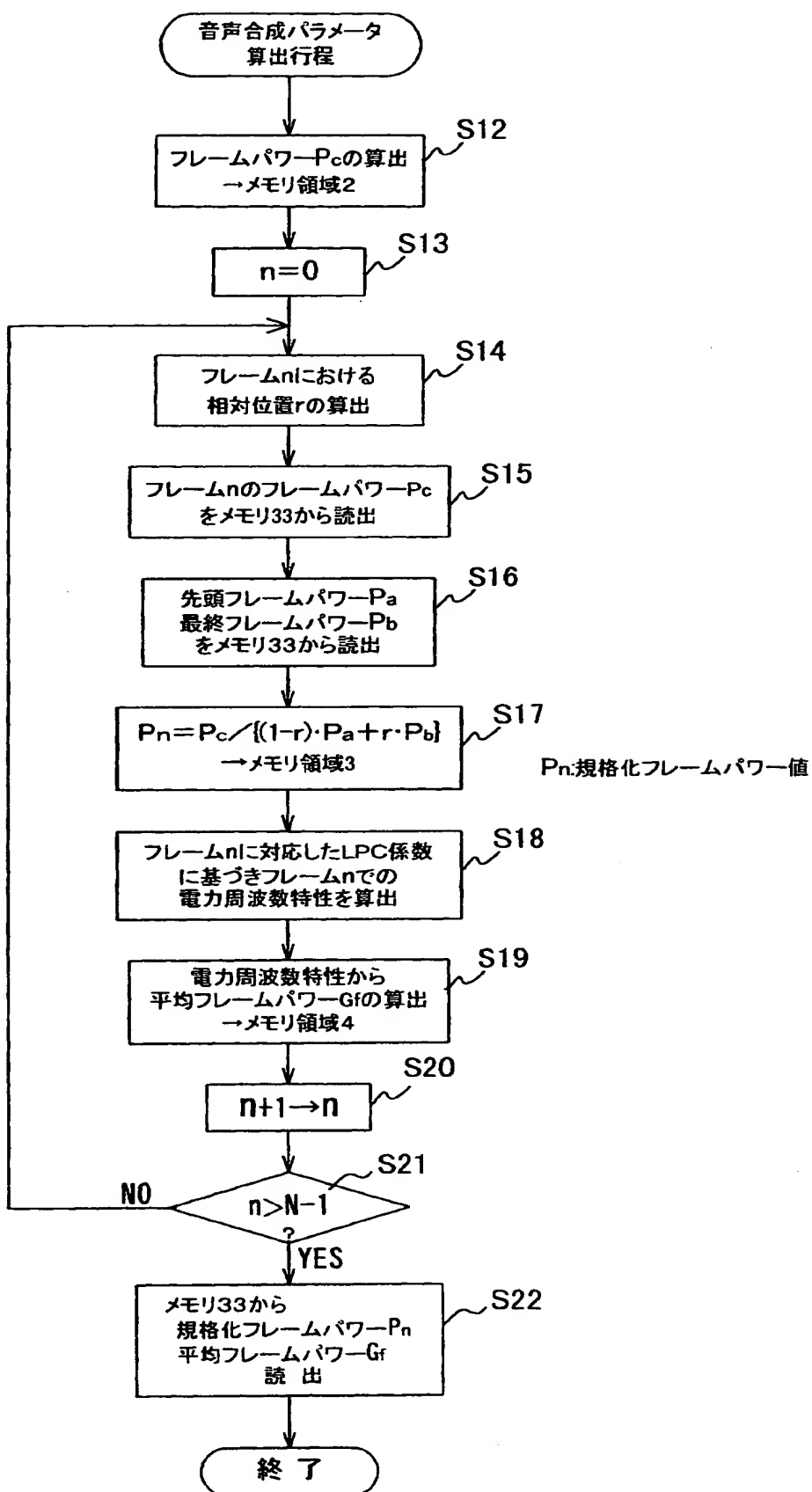
【図 3】



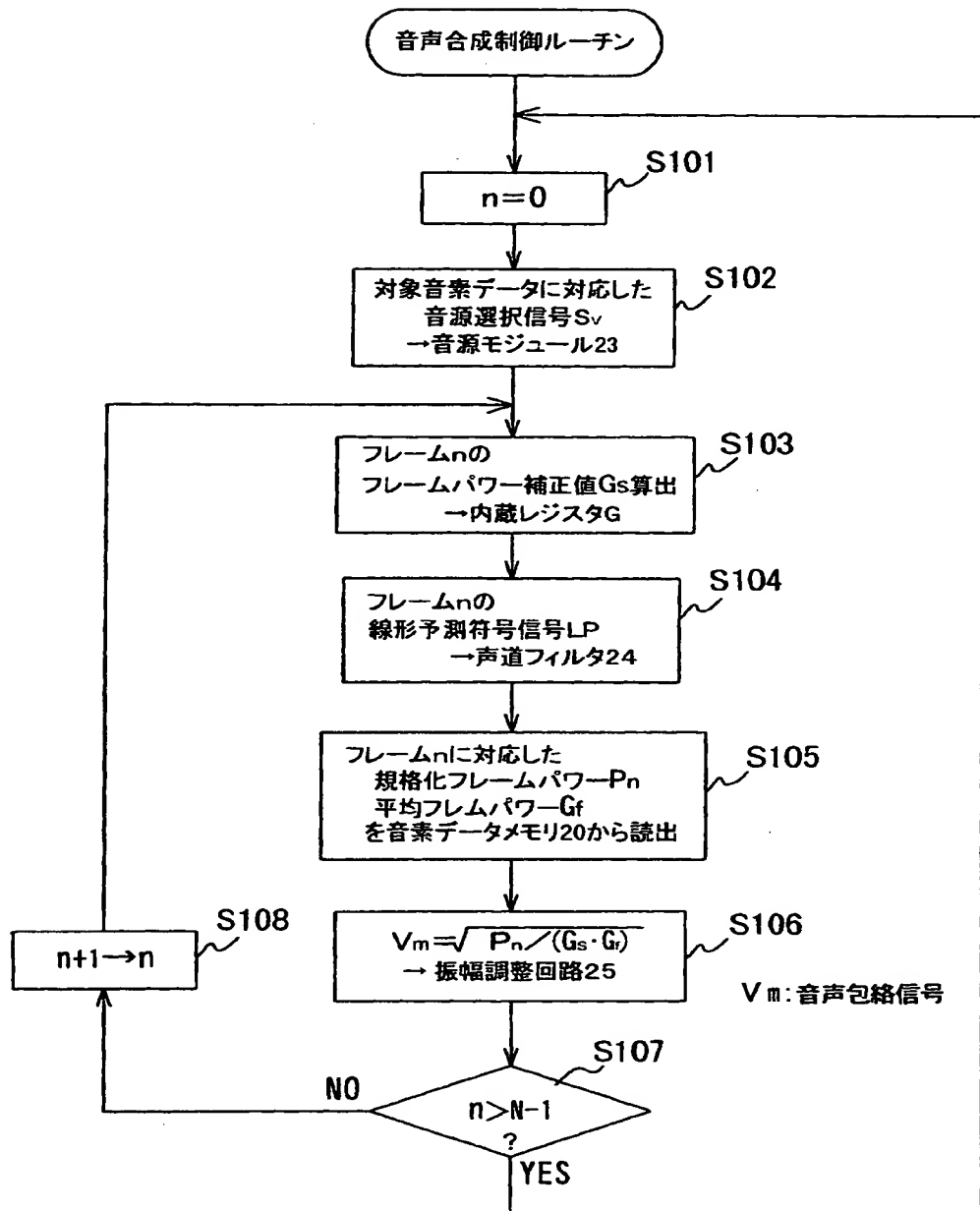
【図 4】



【図 5】



【図 6】



【書類名】 要約書

【要約】

【課題】 自然な合成音声を得られる音声合方法を提供することを目的とする

【解決手段】 音素中の先頭フレーム及び最後尾フレーム各々でのフレームパワー値を所定値に規格化した際における n 番目のフレームでの規格化フレームパワー値を求め、この n 番目のフレームにおける電力周波数特性から所定周波数間隔毎に採取した電力値の平均を平均フレームパワー値として求め、更に、音源からの周波数信号の 1 フレーム期間内での信号レベルの 2 乗和をフレームパワー補正值として夫々求め、これら規格化フレームパワー値、フレームパワー補正值、及び平均フレームパワー値を変数とする関数によって定まるレベルの音声包絡信号に応じて、声道フィルタから出力された音声波形信号の振幅レベルを調整する

【選択図】 図 6

出 願 人 履 歴 情 報

識別番号 [000005016]

1. 変更年月日	1990年 8月31日
[変更理由]	新規登録
住 所	東京都目黒区目黒1丁目4番1号
氏 名	パイオニア株式会社